Enhancing Reinforcement Learning by Offline Pre-Training with Model-Based Augmentation

Alessio Surian

Abstract

Recently, President Obama issued an Executive Order to ensure the United States' leadership in computing. Necessary hardware and software design skills should be introduced into university curricula. Computing has been advanced to High Performance Computing (HPC) throughout the past decades. However, undergraduate students are still lacking of experience in how HPC functions especially in minority-serving institutions, because our current computing curricula do not adequately cover HPC contents. To address this problem, a team of faculty members have obtained external funding supports to improve undergraduate computing education through enhanced courses and research opportunities. The goal is to incorporate HPC concepts and training across the computing curricula in multiple disciplines in order to motivate students' interests in computing and improve their problem-solving skills. This three-year project has already finished the second year of implementation. During the first year, a diverse teaching environment was established, including a HPC cluster and embedded HPC platforms. Both platforms supported students' learning and research in parallel programming, embedded systems design, and data cloud. In the second project year, several courses were revised or developed across three departments: Electrical and Computer Engineering, Computer Science, and Engineering Technology. New course materials integrating the parallel and distributed computing concepts were developed and offered to undergraduate students. Project-based learning was introduced into classroom. More advanced concepts, such as computer vision and machine learning were explored by undergraduate students. At the same time, the research results were disseminated in junior and senior level courses. Faculty members applied effective pedagogy to teach new generation computing. For all the classes involved in this project, student surveys were collected to guide future project implementation. This article shares the current outcomes and findings of the project.

Introduction

Computing technology has advanced to an unbelievable level compared to decades ago. Other than traditional personal computers, there are two types of computing facilities that currently draw much attention from academia as well as industry. One is the computing cluster, the other is the embedded computer. Nowadays if we talk about computer applications, data cloud, virtual reality, and biomedical simulation are often the points of interests. This relies on the fact that Internet and new sensors reach nearly everywhere in our daily lives which leads to an era of big data. To be able to deal with large datasets, the computer industry is moving toward High Performance Computing. HPC system came into being from 1990 and extended the traditional single microprocessor computer to multi-core with parallel programming capability. At the end of the 20th century, computers with thousands of processors were widely used for scientific research. A computer cluster consists of a number of computers to work as a system on computational intensive tasks. Different processors are connected by a network. Shared-memory or distributed memory are the dominate storage types for HPC cluster [1]. The advent of commodity high performance processors, low-latency/high-bandwidth networks, software infrastructure and development tools facilitate the cluster to be widely used for climate modeling, disaster prediction, protein folding, oil and gas industry, and energy research [1, 2]. Currently China's Tianhe [2] is ranking No. 1 among all the super computers based on TOP500 project. Titan (Oak Ridge National Lab) and Sequoia (Lawrence Livermore National Lab) reside in the 2nd and 3rd places respectively [3]. In order to keep the USA's leadership in supercomputers, educators need to provide necessary training to strengthen and diversify the future U.S. workforce in HPC.

Besides the cluster, embedded HPC is also booming quickly. An embedded system performs a specific task and has a computer embedded inside

[4]. Modern embedded systems are usually based on microcontrollers which package processor and peripherals in a single chip. The cell phone as a good example of an embedded system which has been developed to be more portable. With more sophisticated smart phone apps being developed, the smart phones may replace personal computers in the near future. Both the big scale of cluster and the small scale of embedded system bring new implementations which are essential tools to teach new generation computing.

The fast development of computing technologies brings up new challenges to computer education. It has been proved that today's HPC is tomorrow's personal computer. College students, especially students in computing disciplines, need more knowledge of HPC to be able to fulfill tomorrow's workforce needs. How to efficiently teach the young generation computing courses by infusing HPC concepts is critical. Most first tier research institutions utilized their research resources to develop new courses introducing HPC. Some of the successful courses are shared online. But without a local touchable platform, those abstract design concepts are very challenging to be understood by students from underrepresented groups. To address this problem, recently Prairie View A&M University (PVAMU) initiated an NSF funded HBCU Targeted Infusion project to teach undergraduate students HPC through both cluster and embedded platforms. In this three-year project, the research and education foundation was developed in the first year. It just finished the second year's implementation. During the second year, a sequence of courses were revamped and new courses, such as "Parallel Programming" and "Machine Learning," were developed. Several developed or revised courses were offered the first time to undergraduate students in the year of 2014 and 2015. Student surveys were collected to analyze the effectiveness of the class activities and to guide future project implementation. Based on the teaching materials developed in this project, a summer workshop was conducted and project results were disseminated to the public.

Activities and Feedbacks

To enhance computing education, the project activities range from curricula revamping to teaching renovation and out of class undergraduate students research projects. Major activities of this project can be categorized as:

1. Establishing an HPC platform at PVAMU to promote multidisciplinary collaborations for teaching and research.
2. Revamping and developing computing courses across three departments in the College of Engineering at PVAMU to incorporate HPC.
3. Training undergraduate students by involving them in HPC research projects.
4. Enhancing faculty expertise through research and teaching and disseminating results and findings for academic community.

Four faculty members from three departments collaborate in this project. Because computing courses will be taught in three departments, this project directly impacts all the enrolled students in the three departments: Electrical and Computer Engineering, Computer Science, and Engineering Technology. The rest of this section introduces the activities with results from participants' feedback.

Curricula Revamping

The most direct method to enhance computing education is to change the courses offered in computing curricula. At PVAMU, there are three computing related majors: Computer Science (CS), Computer Engineering (CPEG) under Electrical and Computer Engineering department (ECE), and Computer Engineering Technology (CPET) under Engineering Technology department (ET). All the three majors offer different courses on computer hardware and software. The teaching platforms and materials developed can be shared among three departments. The three departments' fall 2015 undergraduate enrollment, categorized by gender and racial/ethnic groups, is listed in Table 1. Of the three programs, female enrollment accounts for 26% and African American students constitute 80% of total enrollment. Together, the three departments constitute 34% of all College of Engineering undergraduate enrollments.

Leveraging a NSF funded Major Research Instrumentation project, an IBM iDataPlex HPC cluster was purchased and resides on campus at the College of Engineering as shown in Figure 1 (a). There are several ongoing projects related to embedded systems, most of them are robotics related. Figure 1 (b) illustrates one embedded HPC platform. It is an unmanned underwater vehicle with ARM technique. These two categories of platforms

can support HPC parallelism productivity using OpenMP, MPI, and Pthread as well as develop computer vision and machine learning courses and research.

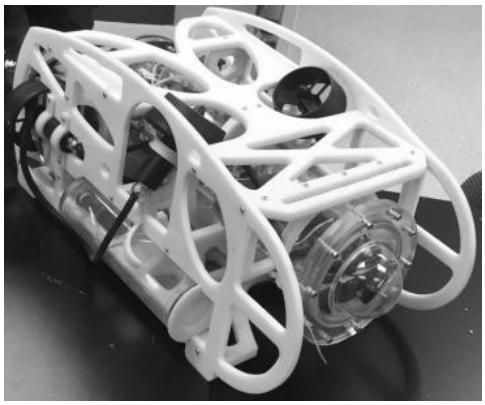
A sequence of courses were redesigned using both embedded and cluster HPC platforms. Two new courses ("Parallel Programming" and "Machine Learning") were developed in 2015. Table 2 lists all the courses designed or revamped in the project, showing the year each course was first offered after revamping/development.

Among courses, ELEG 4253, COMP 4073, and CPET 4053 & 4051 were completed in spring and fall 2014 semesters, so the results were summarized in the previous article [5]. In spring and fall 2015 semesters, ELEG 3073 & 3071, COMP 4073, and CPET 4383 & 4381 were offered the first time after renovation or development. Class surveys and analyses are provided below. All the courses share the same questionnaires as shown in Table 3.

Table 1: Fall 2015 Undergraduate Enrollments of ECE, CS, and ET Departments

Category	ECE	CS	ET
Female	62	37	17
African American	211	101	48
Hispanic	15	9	4
Total	265	127	54

(a) HPC cluster platform.



(b) Embedded HPC platform

Figure 1. Two HPC platforms.

Table 2: Computing Course Involved.

Dept.	**Course**	**New Content**	**Year Offered**
ECE	ELEG 3073 & 3071 Microprocessor Systems Design & lab	Multi-core microprocessor design; Task parallelism; Cloud computing;	2015
	ELEG 4253 Embedded Systems Design	Embedded HPC; FPGA pipelining; Data parallelism; Computer Vision	2014
CS	COMP 4053 Parallel Programming	Parallelism; Cloud computing;	2014
	COMP 4073 Machine Learning	Parallel programming; optimization algorithm, Machine learning	2015
ET	CPET 4383 & 4381 Digital Signal Processing & Lab	Parallelism; Multi-core design; computer vision	2015
	CPET 4053 & 4051 Computer Systems Design & Lab	Multi-core; Cloud computing; Parallelism	2014

Table 3: Pre and Post Survey Questions for Course Involved.

#	Survey Questions	Conduction
1	Consider your level of awareness about High Performance Computing (HPC) both BEFORE and AFTER this class.	*Pre & Post*
2	Consider your level of interest in HPC both BEFORE and AFTER this class.	*Pre & Post*
3	Use the scale to indicate the extent of your gains in understanding of HPC hardware architectures	*Post*
4	Use the scale to indicate the extent of your gains in understanding of HPC system software	*Post*
5	Use the scale to indicate the extent of your gains in understanding of HPC real-world applications	*Post*
6	This class helped me understand the value of the cutting-edge HPC approach to solving problems?	*Post*
7	Knowing more about using HPC techniques will make me more marketable when I graduate?	*Post*
8	I would like HPC to be taught in more classes?	*Post*

ELEG 3073 & 3071 Microprocessors Systems Design & Lab

This lecture and lab bundle introduces the architecture, operation, and applications of microprocessors and microcontrollers. Previously it focused on Intel 8086/8088 series of microprocessors. Recently, PIC microcontroller was also introduced in the teaching materials to add diversity from popular microcontrollers. In the spring 2015 semester, 15% of the class time was spent on HPC related contents. The cluster hardware with examples from the IBM HPC cluster was introduced in class as well as parallel programming and HPC applications. At the end of the class, a survey questionnaire (as shown in Table 3) was given to the whole class of 35 students. Table 4 illustrates the survey results.

COMP 4073 Special Topics-Machine Learning

Based on research findings, a senior level elective COMP 4073 Special Topics-Machine Learning was developed. New teaching materials included both traditional exercises and parallel programming laboratories. The course is new in the CS curriculum. It was offered in spring 2015 for the first time. A total of eight students enrolled in the course. Through the lectures, the basic knowledge of machine learning theory, learning algorithms, and data processing were taught to students. Students showed strong interests in the course contents and gained hands-on skills. By teaching Spark and Scala programming, instructor showed students how to use the latest machine learning tools and how to compare the performance of sequential program and parallel program. Class feedback was very positive (as shown in Table 5).

CPET 4383 & 4381 Digital Signal Processing & Lab

The course and lab introduce basic algorithms for signal processing. After the revision, research projects in computer vision, image acquisition, object detection, etc. were introduced in class. Along with the lectures, hands-on projects allowed students to practice the algorithms and better understand the concepts. All these activities were performed on the proposed hardware and software platforms. Besides the computer vision concepts and algorithms, engineering projects were introduced. The projects enhanced students' understanding of how modern computing techniques are to be explored in real life. Shown in Table 6.

Although three course surveys were obtained from three different departments, they have similar trends. Before the HPC learning modules, most of the students did not know HPC. While after the lectures, most of them showed stronger interests and gained basic skills in using HPC. This project increased undergraduate students' awareness of HPC hardware, software, and applications. Having more HPC modules introduced in class will make students more marketable for the workforce. Based on this successful pilot teaching experience, all the courses will be offered a second round in the next academic year, more exercises will be designed to enhance

Table 4: Pre and post survey results of ELEG 3073 & 3071.

Question #	*Know HPC concepts and applications*	*Know HPC concepts*	*Know only a few about HPC*	*Only heard HPC term*	*Never heard*
1 pre	*2*	*3*	*7*	*12*	*13*
1 post	*13*	*14*	*3*	*5*	*0*
Question #	*Very Interested in HPC*	*Interested in HPC*	*A little interested in HPC*	*Not at all interested in HPC*	*Never heard anything for HPC*
2 pre	*0*	*6*	*13*	*8*	*8*
2 post	*10*	*17*	*4*	*3*	*1*
Question #	*Strongly Agree*	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*
3	*7*	*13*	*11*	*2*	*2*
4	*6*	*11*	*10*	*6*	*2*
5	*4*	*8*	*15*	*5*	*3*
6	*5*	*19*	*8*	*3*	*0*
7	*8*	*18*	*6*	*3*	*0*
8	*5*	*21*	*7*	*2*	*0*

Table 5: Pre and post survey stats of COMP4073-Machine Learning.

Question #	*Know HPC concepts and applications*	*Know HPC concepts*	*Know only a few about HPC*	*Only heard HPC term*	*Never heard*
1 pre	*0*	*2*	*1*	*2*	*2*
1 post	*1*	*6*	*0*	*0*	*0*
Question #	*Very Interested in HPC*	*Interested in HPC*	*A little interested in HPC*	*Not at all interested in HPC*	*Never heard anything for HPC*
2 pre	*1*	*0*	*3*	*0*	*3*
2 post	*3*	*3*	*1*	*1*	*0*
Question #	*A Great Deal*	*A Lot*	*Somewhat*	*A Little*	*Not At All*
3	*0*	*3*	*4*	*0*	*0*
4	*2*	*3*	*1*	*1*	*0*
5	*0*	*4*	*2*	*1*	*0*
Question #	*Strongly Agree*	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*
6	*1*	*5*	*1*	*0*	*0*
7	*2*	*3*	*2*	*0*	*1*
8	*2*	*3*	*2*	*0*	*0*

Table 6: Pre and post survey stats of CPET4383&4381.

Question #	*Know HPC concepts and applications*	*Know HPC concepts*	*Know only a few about HPC*	*Only heard HPC term*	*Never heard*
1 pre	*0*	*2*	*3*	*1*	*2*
1 post	*2*	*3*	*3*	*0*	*0*
Question #	*Very Interested in HPC*	*Interested in HPC*	*A little interested in HPC*	*Not at all interested in HPC*	*Never heard anything for HPC*
2 pre	*0*	*3*	*2*	*1*	*2*
2 post	*0*	*5*	*3*	*0*	*0*
Question #	*A Great Deal*	*A Lot*	*Somewhat*	*A Little*	*Not At All*
3	*2*	*3*	*2*	*1*	*0*
4	*0*	*4*	*3*	*1*	*0*
5	*2*	*2*	*3*	*1*	*0*
Question #	*Strongly Agree*	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*
6	*0*	*6*	*2*	*0*	*0*
7	*0*	*6*	*2*	*0*	*0*
8	*2*	*4*	*2*	*0*	*0*

student understanding of HPC and its applications in real world. Student surveys will be re- designed to collect more in-depth feedback about the new course contents.

Students Training

The curriculum enhancement improves students' skills and confidence, making them better prepared to excel in the computing workforce or continue to graduate school. Each of the four faculty members mentored graduate and undergraduate students on research projects. Four undergraduate students were supported during the past year:

Student A got the fundamentals training of parallel programming, he started working with others in the group to create a web-based parallel programming interface. The goal of the work is to use web-based user-friendly interface to allow students and researchers to learn and develop parallel programs based on MapReduce parallel programming model. The work created a user-friendly and access-anywhere programming environment for the course homework and project in developing parallel programs. The research results were included in new course COMP 4053.

Student B got basic training and worked on installing and configuring of cloud computing server. He is working on HPC, and machine learning (COMP 4073) course material preparation and testing. He is also administering the project website.

Student C was trained on microprocessor knowledge and parallel computing. His research on the IBM cluster was part of the teaching materials for Microprocessor Systems Design (ELEG 3073 & 3071) and Embedded Systems Design (ELEG 4253).

Student D was advised on a parallel image processing algorithm for object detection. He presented his work in a local symposium. His work was introduced in CPET 4383 & 4381 courses.

Summer Workshop

A two-day summer workshop was offered to faculty, staff and students on June 9 and 10, 2015 at PVAMU. Six faculty and research scholars from other universities were invited to give presentations. Eight faculty at PVAMU and thirty students participated in the training and colloquium. One invited talk was focused on the establishment of visualization center for biomedical research. Another invited talk introduced parallel computing education into the entry level CS courses. Workshop survey was conducted by an individual researcher. Data was analyzed and the feedback was positive. Here illustrates sample questions and results:

Question 12: How likely are you going to use the information you obtained from this workshop to your current position?

#	Question	Definitely and Frequently - 5	Definitely and Sometimes - 4	Moderately - 3	Occasionally - 2	Rarely/ Never - 1	Unable to Rate	Total Responses	Mean
1	High Performance Computing and Big Data	15	8	5	1	3	1	33	2.27
2	Usage of HPC Integration in Real World Applications	13	8	6	2	2	2	33	2.45
3	Usage of Information to Impact K-16 Educator in Addressing/Supporting the Computing Industry Workforce Shortage	10	8	7	3	1	4	33	2.82

Question 14: What is your overall perception of the workshop?

#	Question	Significant/ Exceptional - 5	Good - 4	Average - 3	Fair/Marginal - 2	Insignificant/ Poor - 1	Unable to Rate	Total Responses	Mean
1	Overall Workshop Rating	19	10	3	1	0	0	33	4.15

Overall workshop rating was 95 out of 100.

Evidence-based Learning

Besides the technical content revisions, another challenge faced by the educators is how to form a friendly learning environment for computing courses, especially for courses like HPC. This generation grew up with overwhelming fancy computer applications. However their computer related courses are still taught in a traditional way, which leads to high drop-off and failing rate [6-8]. Recent research shows evidence-based teaching strategies [9-11] could be effective to address this problem. Sharing the recent new teaching strategy was brought into the attention from another Department of Education funded project at PVAMU. This strategy was adapted in one of the senior level courses and positive results were obtained from the students' feedback collected at the end of the semester. Authors would like to share the following information.

In fall 2015 semester, a new teaching strategy, sharing the recent news, was implemented in the Embedded System Design course. Students in the class were asked to find recent news about the embedded systems design, parallel computing, and data cloud. In each class meetings, students were given the opportunity to share the recent news with their peers in class. Students introduced the news and discussed them with one another in the first 5 minutes of each class. The goal of the recent news activity is to encourage the students to follow the news on mass media and elsewhere about the course topics. As students get accustomed to reading the news on media about the topic they study in school, it is more likely that they will develop lifelong learning skills [12, 13]. In fact, after their graduation, students will need to be learning about the recent developments in their field from various media including the mass media and news.

Students reported that sharing the recent news activity has provided them with the opportunities to expand their knowledge in the field, including HPC. They believed that reading the recent news and discussing them in class sessions prepared them for future careers because they were able to learn about the most recent project work in the field through the recent news activities.

A student has written that the recent news activity was a helpful method to expand their knowledge and become more prepared to enter the workforce and exposed to different projects. Another student found that sharing the news daily among the class was quite necessary and useful since students were able to know about the new technology that has always being improved by the engineers and scientists working in the field. A student noted that the recent news activity was a very effective method to share the news related to the content of the course, engineering, and as well as technology. She believed that the recent news activity was effective for all students by keeping them abreast on relevant news and topics. She found the value of the recent news activity not only for the senior students but also for all students including freshmen, juniors, and sophomores. By engaging in the recent news

activity, the younger students could keep their attention in the course and gain interest in the material and the industry that will enter after graduation.

Conclusions and Future Work

This article summarized the second year's implementation of an NSF funded project. After the first year's platform establishment, in the second year, all of the proposed courses were able to be developed or revamped and taught for the first time. All the activities were implemented on schedule. Survey results are positive and the project is going toward the correct direction to achieve the following project objectives:

(a) To establish a platform to promote multidisciplinary research collaborations on computing hardware and software design in which students will participate;
(b) To foster undergraduate students' critical thinking skills by involving them in HPC research and in the use of modern computing tools;
(c) To revamp core courses and corresponding labs, and to develop new multidisciplinary courses, to incorporate HPC;
(d) To develop faculty expertise in HPC through research and teaching initiatives;
(e) To disseminate results and findings for academic community and general public aiming at recruiting more students to computing disciplines.

In the future, all the courses will be taught for the second time. Students' feedback will be collected and compared with the first time to get more in-depth analyses to guide future course contents updating. Also the analysis will contribute to the research of developing a more efficient computing education method and pedagogy.

Bibliography

1. P. S. Pacheco, "An Introduction to Parallel Programming," Morgan Kaufman, ISBN: 978-0-12-374260-5.

2. D.A. Bader and R. Pennington, ``Cluster Computing: Applications," *The International Journal of High Performance Computing*, 15(2):181-185, May 2001.

3. Retrieved from http://www.top500.org/lists/ 2015/11 November 2015.

4. J. W. Valvano, "Real-Time Operating Systems for ARM Cortex^{TM} - M Microcontrollers," ISBN: 9781466468863.

5. S. Cui, Y. Wang, L. Li, and L. Huang, "Enhance Computing Curricula with High Performance Computing Teaching and Research," *Proceedings of ASEE 122th Annual Conference & Exposition*, Seattle, WA, June 14-17, 2015.

6. S. Cui, Y. Wang, S. Koay, and Y. Yang, "Revamp Computer Education with Multimedia and Game Technologies," *Proceedings of ASEE 117th Annual Conference & Exposition*, Louisville, Kentucky, June 20-23, 2010.

7. S. Cui, Y. Wang, F. M. Nave, and K. T. Harris, "Teach Computer Techniques through Multimedia," *Proceedings of the 2010 ASEE Northeast Section Conference*, Boston, MA, May 7-8, 2010. http://acamedics.com/proceedings/aseene2010/ papers/facultypapers/80.pdf

8. S. Cui, Y. Yang, A. Lodgher, and K. Phelps, "Strengthen Computing Discipline Recruitment with Educators Collaboration," *Proceedings of the 2011 ASEE Gulf Southwest Annual Conference*, Houston, TX, Mar. 9-10, 2011.

9. A. Forte and M. Guzdial. "Computers for Communication, Not Calculation: Media as a Motivation and Context for Learning," *Proceedings of 37th Hawaiian International Conference of Systems Sciences*, Big Island, Hawaii, January 2004.

10. S. Bermejo, “Cooperative Electronic Learning in Virtual Laboratories Through Forums,” IEEE Transactions on Education, vol. 48, No. 1, February 2005, pp. 140-149.

11. D. Laurillard, *Rethinking University Teaching: A Conversational Framework for the Effective use of Learning Technologies*, 2nd ed. London, U.K.: Routledge/Falmer, 2002, pp. 11–24.

12. B. Yalvac, C. Ehlig-Economides, L. A. Brooks, and D. Smith. “An international, interdisciplinary, and dynamic approach to teaching energy utilization and sustainability,” *Proceedings of the International Symposium on Engineering Education and Educational Technologies (EEET).* Orlando, FL, July, 2009.

13. [13] B. Yalvac, L. A. Brooks, and C. Ehlig-Economides. “Research on the evolution of college instructors’ perspectives of teaching and learning,” *Proceedings of the American Society for Engineering Education (ASEE) Annual Conference and Exposition*, Pittsburgh, PA, June, 2008